

# COVID-19 Case Surveillance Public Use Data with Geography Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value “Unknown” when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value “Missing” when jurisdictions do not provide a value, and the value “NA” when the value is suppressed as part of privacy protections.

Dataset version: 2021-09-28

Total records in dataset: 32,806,678

## Quick Summary

	all_fields	quasi_fields
total_fields	19	8
total_records	32,806,678	32,806,678
total_cells	623,326,882	262,453,424
missing_fields	132,120,999	16,067,349
missing_pct	21%	6%
complete_fields	491,205,883	246,386,075
complete_pct	79%	94%
unknown_fields	29,485,079	14,998,072
unknown_pct	5%	6%
suppressed_fields	19,400,339	17,524,146
suppressed_pct	3%	7%
available_fields	442,320,465	213,863,857
available_pct	71%	81%

## Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
case_month	4	0.0%	0	0.0%
res_state	678	0.0%	0	0.0%
res_county	1,875,515	5.7%	0	0.0%
age_group	377,973	1.2%	323,645	1.0%
sex	1,020,368	3.1%	38,429	0.1%
race	5,846,875	17.8%	2,555,878	7.8%
ethnicity	6,806,875	20.7%	1,897,646	5.8%
death_yn	1,595,858	4.9%	11,251,751	34.3%
records_with_any_field	9,202,923	28.1%	12,469,753	38.0%